ORTHOGONAL TENSOR DECOMPOSITIONS*

TAMARA G. KOLDA†

Abstract. We explore the orthogonal decomposition of tensors (also known as multi-dimensional arrays or n-way arrays) using two different definitions of orthogonality. We present numerous examples to illustrate the difficulties in understanding such decompositions. We conclude with a counterexample to a tensor extension of the Eckart-Young SVD approximation theorem by Leibovici and Sabatier [*Linear Algebra Appl.* 269(1998):307–329].

Key words. tensor decomposition, singular value decomposition, principal components analysis, multidimensional arrays

AMS subject classifications. 15A69, 49M27, 62H25

1. Introduction.

The problem of decomposing tensors (also called n-way arrays or multidimensional arrays) is approached in a variety of ways by extending the singular value decomposition (SVD), principal components analysis (PCA), and other methods to higher orders; see, e.g., [1, 3, 9, 10, 11, 12, 13, 14, 15]. Tensor decompositions are most often used for multimode statistical analysis and clustering, but may also be used for compression of multidimensional arrays in ways similar to using a low-rank SVD for matrix compression. For example, color images are often stored as a sequence of RGB triplets, i.e., as separate red, green and blue overlays. An $m \times n$ pixel RGB image is represented by an $m \times n \times 3$ array, and a collection of p such images is an $m \times n \times 3 \times p$ array and can be compressed by a low-rank approximation.

The notation and basic properties of tensors are set forth in §2. Several definitions of orthogonality and several rank orthogonal decompositions for tensors are given in §3. Computational issues for orthogonal decompositions are discussed in §4. Finally in §5, we present a counterexample to Leibovici and Sabatier's extension to tensors of the well-known Eckart-Young SVD approximation theorem [13].

2. Tensors.

Let A be an $m_1 \times m_2 \times \cdots \times m_n$ *tensor* over $\mathbb{R}$. The *order* of A is n. The jth *dimension* of A is m_j. An element of A is specified as

$$A_{i_1 i_2 \cdots i_n},$$

where $i_j \in \{1, 2, \ldots, m_j\}$ for $j = 1, \ldots, n$. The set of all tensors of size $m_1 \times m_2 \times \cdots \times m_n$ is denoted by $\mathcal{T}(m_1, m_2, \ldots, m_n)$. The shorthand $\mathcal{T}_n$ may be used when only the order needs to be specified, or just $\mathcal{T}$ may be used when the order and dimensions are unambiguous.

Let $A, B \in \mathcal{T}(m_1, m_2, \ldots, m_n)$. The *inner product*[1] of A and B is defined as

$$A \cdot B \equiv \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} A_{i_1 i_2 \cdots i_n} B_{i_1 i_2 \cdots i_n}.$$

* This work was supported by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy, under contracts DE–AC05–96OR22464 with Lockheed Martin Energy Research Corporation and DE-AC04-94AL85000 with Sandia Corporation.

†Computational Science and Mathematics Research Department, Sandia National Laboratories, Livermore, CA 94551–9217, tgkolda@sandia.gov.

[1]In [13], the term is "contracted product" and the notation is $\langle A, B\rangle$.

Correspondingly, the *norm* of A, $\|A\|$, is defined as

$$\|A\|^2 \equiv A \cdot A = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} A_{i_1 i_2 \cdots i_n}^2.$$

We say A is a *unit* tensor if $\|A\| = 1$.

Example 2.1. Let $x, y \in \mathcal{T}(m)$; that is, x, y are vectors in $\mathbb{R}^m$. Then $x \cdot y = x^T y$ where the superscript T denotes transpose. □

A *decomposed tensor* is a tensor $U \in \mathcal{T}(m_1, m_2, \ldots, m_n)$ that can be written as

$$U = u^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(n)}, \tag{2.1}$$

where $\otimes$ denotes the outer product and each $u^{(j)} \in \mathbb{R}^{m_j}$ for $j = 1, \ldots, n$. The vectors $u^{(j)}$ are called the *components* of U. In this case,

$$U_{i_1 i_2 \cdots i_n} = u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_n}^{(n)}.$$

A decomposed tensor is a tensor of rank one for all the definitions of rank that we present in the next section. Decomposed tensors form the building blocks for tensor decompositions. The set of all decomposed tensors of size $m_1 \times m_2 \times \cdots \times m_n$ is denoted by $\mathcal{D}(m_1, m_2, \ldots, m_n)$ with shorthands analogous to $\mathcal{T}$.

LEMMA 2.2. *Let $U, V \in \mathcal{D}$ where U is defined as in (2.1) and V is defined by*

$$V = v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(n)}. \tag{2.2}$$

Then

$$\text{(a)} \quad U \cdot V = \prod_{j=1}^{n} u^{(j)} \cdot v^{(j)}, \qquad \text{(b)} \quad \|U\| = \prod_{j=1}^{n} \|u^{(j)}\|_2,$$

and, (c) $U + V \in \mathcal{D}$ if and only if all but at most one of the components of U and V are equal (within a scalar multiple).

Proof. Items (a) and (b) follow directly from the definitions. For item (c), consider $U, V \in \mathcal{D}$ such that $n-1$ components are equal, i.e., $u^{(i)} = v^{(i)}$ for $i = 2, \ldots, n$. Then $W \equiv U + V$ can be written as

$$W = w^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(n)},$$

where $w^{(1)} = u^{(1)} + v^{(1)}$, so the "if" statement of (c) is true. Next we show the "only if" statement of (c). First consider the special case where $n = 2$, $m_1 = m_2 = 2$,

$$U \equiv \begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} c \\ d \end{bmatrix}, V \equiv \begin{bmatrix} e \\ f \end{bmatrix} \otimes \begin{bmatrix} g \\ h \end{bmatrix},$$

and $W \equiv U + V \in \mathcal{D}$. Since $W \in \mathcal{D}$, we can write it as

$$W \equiv \begin{bmatrix} p \\ q \end{bmatrix} \otimes \begin{bmatrix} r \\ s \end{bmatrix}.$$

Then, we have

$$pr = ac + eg, \tag{2.3}$$
$$ps = ad + eh, \tag{2.4}$$
$$qr = bc + fg, \tag{2.5}$$
$$qs = bd + fh. \tag{2.6}$$

Dividing (2.3) by (2.5) and (2.4) by (2.6) yields two ratios for p/q, and setting those equals gives

$$\frac{ac+eg}{bd+fh} = \frac{bc+fg}{ad+eh}. \tag{2.7}$$

Cross-multiplying and simplifying (2.7) finally yields

$$(af-be)(ch-dg) = 0.$$

In other words, either $u^{(1)} = v^{(1)}$ or $u^{(2)} = v^{(2)}$ (within a scalar multiple). So, all but at most one of the components of U and V must match if $W \in \mathcal{D}$. This argument can be extended to arbitrary n and m_j. $\square$

We have shown that for two decomposed tensors to be combined to one decomposed tensor, they must match in all but at most one component. The same is not necessarily true, however, when combining three or more decomposed tensors, as shown in the next example.

Example 2.3. Consider the following example. Let $a, b \in \mathbb{R}^m$ with $a \perp b$ and $\|a\| = \|b\| = 1$. Define $c = \frac{1}{\sqrt{2}}(a+b)$, and

$$U_1 = a \otimes a \otimes a, \quad U_2 = a \otimes b \otimes c, \quad U_3 = a \otimes c \otimes b.$$

Then the sum of theses three decomposed tensors can be rewritten as the sum of two despite the fact that they only match in one component:

$$U_1 + U_2 + U_3 = \sqrt{\frac{3}{2}}\,(V_1 + V_2)\,,$$

where

$$V_1 = a \otimes d \otimes a, \quad V_2 = a \otimes e \otimes b,$$

with

$$d = \sqrt{\frac{2}{3}}\,a + \sqrt{\frac{1}{3}}\,b, \quad e = \sqrt{\frac{2}{3}}\,c + \sqrt{\frac{1}{3}}\,b.$$

This is the result of splitting U_2 into two pieces based on the third component. $\square$

We may also operate on tensors of different sizes. Specifically, tensors of different orders may be multiplied as follows. Suppose $C \in \mathcal{T}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_n)$ is a tensor of order $n-1$ (note that m_j is missing). Then the *contracted product*[2] of A and C is a vector of length m_j, and its i_jth ($1 \le i_j \le m_j$) element is defined as

$$\langle A \cdot C \rangle^{(j)}_{i_j} \equiv \sum_{i_1=1}^{m_1} \cdots \sum_{i_{j-1}=1}^{m_{j-1}} \sum_{i_{j+1}=1}^{m_{j+1}} \cdots \sum_{i_n=1}^{m_n} A_{i_1 \cdots i_{j-1} i_j i_{j+1} \cdots i_n} C_{i_1 \cdots i_{j-1} i_{j+1} \cdots i_n}.$$

Note that the superscript on the bracketed product indicates which dimension is missing in the lower-order tensor C.

Example 2.4. Suppose $A \in \mathcal{T}(m_1, m_2)$ is a tensor of order two, i.e., A is a matrix. If $b \in \mathcal{T}(m_1)$, then $\langle A \cdot b \rangle^{(2)} = A^T b$ in matrix notation. Similarly, if $c \in \mathcal{T}(m_2)$, then $\langle A \cdot c \rangle^{(1)} = Ac$. $\square$

[2] In [13], the notation $A \mathbin{..} C$ is used for contracted products.

LEMMA 2.5. *Let $U \in \mathcal{D}$ as defined in (2.1) and $A \in \mathcal{T}$. Then*

$$A \cdot U = \left\langle A \cdot u^{(1)} \otimes \cdots \otimes u^{(j-1)} \otimes u^{(j+1)} \otimes \cdots \otimes u^{(n)} \right\rangle^{(j)} \cdot u^{(j)}.$$

Proof. Follows from the definitions. □

3. Orthogonal Rank Decompositions.

3.1. Notions of Orthogonality. Let $U, V \in \mathcal{D}$ be defined as in (2.1) and (2.2) respectively. Without loss of generality, we assume $\|U\| = \|V\| = 1$ and that the components are unit vectors. We say that U and V are *orthogonal* ($U \perp V$) if

$$U \cdot V = \prod_{j=1}^{n} u^{(j)} \cdot v^{(j)} = 0.$$

We say that U and V are *completely orthogonal* ($U \perp_{\mathrm{c}} V$) if for every $j = 1. \ldots, n$,

$$u^{(j)} \perp v^{(j)}.$$

We say that U and V are *strongly orthogonal* ($U \perp_{\mathrm{s}} V$) if $U \perp V$ and for every $j = 1, \ldots, n$,

$$u^{(j)} = \pm\, v^{(j)} \text{ or } u^{(j)} \perp v^{(j)}.$$

From the definition of strong orthogonality, it follows that at least one pair must satisfy $u^{(j)} \perp v^{(j)}$ since we require $U \perp V$. Note that we could write $u^{(j)} = \pm\, v^{(j)}$ more generally as $u^{(j)} = \lambda_j\, v^{(j)}$ for some $\lambda_j \neq 0$, which is useful when $\|U\| \neq \|V\|$.

The relationship between the different orthogonality definitions is given in the following lemma.

LEMMA 3.1. *Let the decomposed tensors U and V of order n be defined as in (2.1) and (2.2) respectively. Then*

$$U \perp_{\mathrm{c}} V \Rightarrow U \perp_{\mathrm{s}} V \Rightarrow U \perp V.$$

3.2. Rank Decompositions. Our goal is to express a tensor $A \in \mathcal{T}$ as a weighted sum of decomposed tensors,

$$A = \sum_{i=1}^{r} \sigma_i U_i, \tag{3.1}$$

where $\sigma_i > 0$ for $i = 1, \ldots, r$ and each $U_i \in \mathcal{D}$ and $\|U_i\| = 1$ for $i = 1, \ldots, r$.

- The *rank* of A, denoted $\operatorname{rank}(A)$, is defined to be the minimal r such that A can be expressed as in (3.1). The decomposition is called the *rank decomposition.*
- The *orthogonal rank* of A, denoted $\operatorname{rank}_{\perp}(A)$, is defined to be the minimal r such that A can be expressed as in (3.1) and $U_i \perp U_j$ for all $i \neq j$. The decomposition is called the *orthogonal rank decomposition.*
- The *strong orthogonal rank* of A, denoted $\operatorname{rank}_{\perp_{\mathrm{s}}}(A)$, is defined to be the minimal r such that A can be expressed as in (3.1) and $U_i \perp_{\mathrm{s}} U_j$ for all $i \neq j$. The decomposition is called the *strong orthogonal rank decomposition.*[3]

[3] In [13], the terms "free orthogonal rank" and "free rank decomposition" are used rather than "strong orthogonal rank" and "strong orthogonal rank decomposition".

As reported in [13], the definition of rank is due to Kruskal, although it was proposed even earlier by Strassen and others (see [11] and references therein), and the definitions of orthogonal and strong orthogonal rank is due to Franc [7]. The general *decomposition*, *orthogonal decomposition*, and *strong orthogonal decomposition* satisfy the orthogonality constraints (if any) but are not necessarily minimal in terms of r. For matrices, all three rank decompositions are equivalent to the SVD.

LEMMA 3.2. *The rank, orthogonal rank, and strong orthogonal rank decomposition are each equivalent to the SVD for tensors of order two.*

Proof. This follows from the properties of the SVD (c.f., [8]). □

In our discussion of rank decomposition, we did not present a *completely orthogonal decomposition.* In fact, we are not in general guaranteed that such a decomposition can be found, as we discuss later in this section.

A slightly different notion of rank that depends on special orthogonal decomposition is the *combinatorial orthogonal rank*, denoted $\operatorname{rank}_{\perp_t}(A)$. It is defined as the minimal r such that A can be written as

$$\sum_{i_1=1}^{r}\sum_{i_2=1}^{r}\cdots\sum_{i_n=1}^{r}\sigma_{i_1 i_2\cdots i_n}\, u_{i_1}^{(1)}\otimes u_{i_2}^{(2)}\otimes\cdots u_{i_n}^{(n)}, \tag{3.2}$$

where $\sigma_{i_1 i_2\cdots i_n} > 0$; $u_i^{(j)} \in \mathbb{R}^{m_j}$ with $\|u_i^{(j)}\| = 1$ for $1 \le i \le r$ and $1 \le j \le n$; and further, $u_{i_1}^{(j)} \perp u_{i_2}^{(j)}$ for all $i_1 \ne i_2$, $1 \le i_2, i_2 \le r$, $1 \le j \le n$. Equivalently, let

$$U_i = u_i^{(1)}\otimes u_i^{(2)}\otimes\cdots\otimes u_i^{(n)},$$

and require $U_{i_1} \perp_c U_{i_2}$ for all $i_1 \ne i_2$, $1 \le i_1, i_2 \le r$, and $\|U_i\| = 1$, $1 \le i \le r$. In other words, the decomposition (3.2) is the result of combining the components of the U_i's in every possible way and is called the *combinatorial orthogonal rank decomposition.* In this case, there are r^n scalar multiples (i.e., σ-values) that are involved rather than just r as in the other decompositions. This is the Tucker decomposition with orthogonality constraints [14], hence the subscript in the notation. Note that the SVD of a matrix is a combinatorial orthogonal rank decomposition, but the reverse is not necessarily true.

Now we consider several examples that illustrate that the rank decompositions are not necessarily unique.

Example 3.3. Let $a, b \in \mathbb{R}^m$ with $a \perp b$ and $\|a\| = \|b\| = 1$, and let $\sigma_1 > \sigma_2 > \sigma_3 > 0$. Define $A \in \mathcal{T}(m, m, m)$ as

$$A = \sigma_1 \underbrace{a\otimes b\otimes b}_{U_1} + \sigma_2 \underbrace{b\otimes b\otimes b}_{U_2} + \sigma_3 \underbrace{a\otimes a\otimes a}_{U_3}. \tag{3.3}$$

Note that $U_i \perp_s U_j$ for all $i \ne j$, so (3.3) is a strong orthogonal decomposition of A. Furthermore, A cannot be expressed as the sum of fewer weighted strong orthogonal decomposed tensors, so the strong orthogonal rank of A is three. Observe that A can also be expressed as

$$A = \hat\sigma_1 \underbrace{\hat a\otimes b\otimes b}_{\hat U_1} + \hat\sigma_2 \underbrace{\hat a\otimes a\otimes a}_{\hat U_2} + \hat\sigma_3 \underbrace{\hat b\otimes a\otimes a}_{\hat U_3}, \tag{3.4}$$

where

$$\hat\sigma_1 = \sqrt{\sigma_1^2+\sigma_2^2}\ , \quad \hat\sigma_2 = \frac{\sigma_1\,\sigma_3}{\hat\sigma_1}\ , \quad \hat\sigma_3 = \frac{\sigma_2\,\sigma_3}{\hat\sigma_1}\ ,$$

$$\hat{a} = \frac{\sigma_1\ a + \sigma_2\ b}{\hat{\sigma}_1}\ , \quad \text{and} \quad \hat{b} = \frac{\sigma_2\ a - \sigma_1\ b}{\hat{\sigma}_1}\ .$$

Since $\hat{a} \perp \hat{b}$, we have $\hat{U}_i \perp_s \hat{U}_j$ for all $i \neq j$. Therefore (3.4) is also a strong orthogonal rank decomposition of A, and so the strong orthogonal rank decomposition is not unique. □

Example 3.4. Consider the tensor A as defined by (3.3); A can also be written as

$$A = \bar{\sigma}\bar{U} + \sigma_3 U_3, \tag{3.5}$$

where

$$\bar{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad \bar{U} = \frac{\sigma_1\ a + \sigma_2\ b}{\bar{\sigma}} \otimes b \otimes b.$$

Observe that $\bar{U} \perp U_3$; in fact, (3.5) is an orthogonal rank decomposition of A, and therefore the orthogonal rank of A is two. An alternative orthogonal rank decomposition of A is given by

$$A = \tilde{\sigma}\ \tilde{U} + \sigma_2 U_2, \tag{3.6}$$

where

$$\tilde{\sigma} = \sqrt{\sigma_1^2 + \sigma_3^2} \quad \text{and} \quad \tilde{U} = a \otimes \frac{\sigma_1\ b + \sigma_3\ a}{\tilde{\sigma}} \otimes b.$$

Note that $\tilde{U} \perp \hat{U}_2$, so (3.6) is also an orthogonal rank decomposition of A and the orthogonal rank decomposition is not unique. □

LEMMA 3.5. *Neither the orthogonal rank, strong orthogonal rank, nor combinatorial orthogonal rank decomposition is unique.*

Proof. See Examples 3.3 and 3.4. □

Although the singular value decomposition for matrices in known to be unique up to rotation [8], the rank tensor decompositions are not. This is an important difference which we return to later in this section.

Example 3.6. We show how to 'orthogonalize' a tensor in a relatively simple situation. Suppose that we have an order three tensor $A \in \mathcal{T}(m_1, m_2, m_3)$ defined as follows:

$$A = \sigma_1 U + \sigma_2 V,$$

where $\sigma_1 \geq \sigma_2$ and,

$$U = u^{(1)} \otimes u^{(2)} \otimes u^{(3)},$$
$$V = v^{(1)} \otimes v^{(2)} \otimes v^{(3)},$$

with $u^{(i)}, v^{(i)}$ unequal, non-orthogonal unit vectors in $\mathbb{R}^{m_i}$ for $i = 1, 2, 3$.

For $i = 1, 2, 3$, we can decompose $v^{(i)}$ as

$$v^{(i)} = \alpha^{(i)} u^{(i)} + \hat{\alpha}^{(i)} \hat{u}^{(i)},$$

where

$$\alpha^{(i)} = v^{(i)} \cdot u^{(i)},$$
$$\hat{\alpha}^{(i)} = \|v^{(i)} - \alpha^{(i)} u^{(i)}\|, \text{ and}$$
$$\hat{u}^{(i)} = (v^{(i)} - \alpha^{(i)} u^{(i)}) / \hat{\alpha}^{(i)}.$$

Then, we can rewrite A as

$$
\begin{array}{rcll}
A & = & (\sigma_1 + \sigma_2\, \alpha^{(1)}\alpha^{(2)}\alpha^{(3)}) & u^{(1)} \otimes u^{(2)} \otimes u^{(3)} \\
& + & \sigma_2\, \alpha^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)} & u^{(1)} \otimes u^{(2)} \otimes \hat{u}^{(3)} \\
& + & \sigma_2\, \alpha^{(1)}\hat{\alpha}^{(2)}\alpha^{(3)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes u^{(3)} \\
& + & \sigma_2\, \alpha^{(1)}\hat{\alpha}^{(2)}\hat{\alpha}^{(3)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes \hat{u}^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\alpha^{(2)}\alpha^{(3)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes u^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes \hat{u}^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\hat{\alpha}^{(2)}\alpha^{(3)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes u^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\hat{\alpha}^{(2)}\hat{\alpha}^{(3)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes \hat{u}^{(3)}.
\end{array} \tag{3.7}
$$

Equation (3.7) shows that $\operatorname{rank}_{\perp_s}(A) \leq 8$. Because of the way U and V were chosen (components neither equal nor orthogonal), equation (3.7) is a strong orthogonal rank decomposition of A, and $\operatorname{rank}_{\perp_s}(A) = 8$. (From Equation (3.7), we can also deduce that $\operatorname{rank}_{\perp_t}(A) = 2$.) This is not, however, an orthogonal rank decomposition. Combining each pair of lines in (3.7), we get

$$
\begin{array}{rcll}
A & = & \sqrt{\gamma^2 + \hat{\gamma}^2} & u^{(1)} \otimes u^{(2)} \otimes (\gamma u^{(3)} + \hat{\gamma}\hat{u}^{(3)})/\sqrt{\gamma^2 + \hat{\gamma}^2} \\
& + & \sigma_2\, \alpha^{(1)}\hat{\alpha}^{(2)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\alpha^{(2)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes v^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)}\hat{\alpha}^{(2)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)}.
\end{array} \tag{3.8}
$$

where

$$\gamma = \sigma_1 + \sigma_2\, \alpha^{(1)}\alpha^{(2)}\alpha^{(3)} \text{ and } \hat{\gamma} = \sigma_2\, \alpha^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)}.$$

Finally, combining the last two lines of (3.8), we arrive at an orthogonal rank decomposition,

$$
\begin{array}{rcll}
A & = & \sqrt{\gamma^2 + \hat{\gamma}^2} & u^{(1)} \otimes u^{(2)} \otimes (\gamma u^{(3)} + \hat{\gamma}\hat{u}^{(3)})/\sqrt{\gamma^2 + \hat{\gamma}^2} \\
& + & \sigma_2\, \alpha^{(1)}\hat{\alpha}^{(2)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)} \\
& + & \sigma_2\, \hat{\alpha}^{(1)} & \hat{u}^{(1)} \otimes v^{(2)} \otimes v^{(3)},
\end{array}
$$

so $\operatorname{rank}_{\perp}(A) = 3$. Note that combining vectors from (3.7) in different order would have resulted in a different orthogonal rank decomposition. ▯

We now see some relationship between the different ranks, stated formally in the next theorem.

Theorem 3.7 ([13]). *For a given tensor A,*

$$\operatorname{rank}(A) \leq \operatorname{rank}_{\perp}(A) \leq \operatorname{rank}_{\perp_s}(A). \tag{3.9}$$

Further, for any order $n > 2$, there exists $A \in \mathcal{T}_n$ such that strict inequality holds.

Proof. The first part follows from Lemma 3.1. An example of strict inequality for a tensor of order three ($n = 3$) is given in Example 3.6, and that example can be generalized to any order. ▯

For a matrix, all four definitions of tensor rank reduce to the standard definition of matrix rank.

Corollary 3.8 ([13]). *For any $A \in \mathcal{T}_2$,*

$$\operatorname{rank}(A) = \operatorname{rank}_{\perp}(A) = \operatorname{rank}_{\perp_s}(A) = \operatorname{rank}_{\perp_t}(A).$$

Proof. This follows from Lemma 3.2. □

Earlier we mentioned the notion of a completely orthogonal decomposition; this corresponds to a combinatorial orthogonal decomposition in which only the diagonal elements ($\sigma_{ii\cdots i}$) are nonzero; and so, in general, tensors cannot be *diagonalized.* A similar observation was made by Denis and Dhorne [4]. When a tensor can be diagonalized, all the ranks are equal.

COROLLARY 3.9 ([13]). *For any order $n > 2$, there exists $A \in \mathcal{T}_n$ such that A cannot be decomposed as the weighted sum of completely orthogonal tensors. If a tensor can be decomposed as the weighted sum of completely orthogonal decomposed tensors, then equality holds in (3.9).*

Proof. See the construction of the decompositions of A in Example 3.6 to prove the first statement. The second statement follows intuitively from the fact that each subspace has dimension r, and the rank of the tensor cannot be less than the smallest-dimensional subspace. □

Franc [6] made similar observations to Theorem 3.7 and Corollary 3.9. Matrices (i.e., tensors of order two) are special cases that always have a completely orthogonal decomposition, as follows from Corollaries 3.8 and 3.9.

We now return to the concept of uniqueness in the rank decomposition. We have several examples illustrating that the strong orthogonal rank and orthogonal rank decompositions are not unique. A partial 'fix' for lack of uniqueness is the following. Without loss of generality, assume that the σ_i's in (3.1) are always ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then define the *unique* (strong) orthogonal rank decomposition to be the (strong) orthogonal rank decomposition that has the largest possible σ_1, and given that choice for σ_1, has the largest possible σ_2, and so forth. This decomposition is unique in the sense that the weights are unique. The unit decomposed tensors are unique if and only if no two σ_i's are equal, similar to the fact that the SVD is unique up to rotation. A *unique* combinatorial orthogonal rank decomposition can be defined in a more complicated way by sequentially choosing each U_k so that

$$\sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_n=1}^{k} \sigma^2_{i_1 i_2 \cdots i_n},$$

is maximized.

Example 3.10. In Example 3.3, the unique strong orthogonal rank decomposition is given by (3.4). Similarly, in Example 3.4, the unique orthogonal rank decomposition is given by (3.5). □

4. Greedy Tensor Decompositions. We now consider the computation of an orthogonal decomposition and present a method for generating a *greedy orthogonal decomposition.* Our goal is to compute a sequence (for $p = 1, 2, \ldots$) of weighted decomposed tensors such that

$$A = \sum_{i=1}^{p} \sigma_i U_i,$$

where $U_i \perp U_j$ for all $i \neq j$ and $\|U_i\| = 1$ for all i. We call this the greedy orthogonal decomposition because the $\{\sigma, U\}$ pairs are computed iteratively. We do not yet make any claims as to whether or not this greedy orthogonal decomposition yields an orthogonal rank decomposition.

In the greedy orthogonal decomposition, define the kth residual tensor as

$$R_k \equiv A - \sum_{i=1}^{k} \sigma_i U_i,$$

with $R_0 = A$, and let the set of tensors $\mathcal{U}_k$ be defined as

$$\mathcal{U}_k = \{U_1, U_2, \ldots, U_k\},$$

with $\mathcal{U}_0 = \emptyset$. Our goal is to find the best rank-1 approximation to the current residual subject to orthogonality constraints; that is, we wish to solve

$$\min \ f_k(\sigma, U) \equiv \|R_k - \sigma U\|^2, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp \mathcal{U}_k.$$

We can rewrite f_k as

$$f_k(\sigma, U) = \|R_k\|^2 - 2\sigma R_k \cdot U + \sigma^2 \|U\|^2.$$

At the solution, we have

$$\frac{\partial f_k}{\partial \sigma} \equiv -2R_k \cdot U + 2\sigma \|U\|^2 = 0,$$

so we can solve for σ and conclude that minimizing f_k is the same as solving

$$\max \ R_k \cdot U, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp \mathcal{U}_k. \tag{4.1}$$

We define U_{k+1} to be the solution of (4.1), and let $\sigma_{k+1} = R_k \cdot U_{k+1}$. We repeat the process until $R_{k+1} = 0$.

A *greedy strong orthogonal decomposition* can be similarly described, and reduces to solving

$$\max \ R_k \cdot U, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp_{\mathrm{s}} \mathcal{U}_k, \tag{4.2}$$

at each iteration. Likewise, We may also construct a sort of greedy approach for the combinatorial orthogonal decomposition.

LEMMA 4.1. *The greedy orthogonal, strong orthogonal, and combinatorial decompositions are finite.*

Proof. This is a consequence of the fact that there are at most $M = \prod_{j=1}^{n} m_j$ orthogonal or strong orthogonal decomposed tensors. ∎

Solving (4.1) or (4.2) is a very challenging task. For example, in order to solve (4.1), we might use an *alternating least squares* (ALS) approach as follows. For $\ell = 1, \ldots, n$, fix all components of U but the ℓth, and solve

$$\max \ s \cdot u^{(\ell)}, \ \text{ s.t. } \ \|U\| = 1, \ \ U \perp \mathcal{U}_k$$

where

$$s = \left\langle R_k \cdot u^{(1)} \otimes \cdots \otimes u^{(\ell-1)} \otimes u^{(\ell+1)} \otimes \cdots \otimes u^{(n)} \right\rangle^{(\ell)}.$$

The difficulty with this approach is in enforcing the constraints.

Zhang and Golub [15] explore various computational techniques when the tensor has a completely orthogonal decomposition, in which case the problem is much simpler. In [13], the RPVSCC method uses ALS to find the *modes*, i.e., the completely orthogonal decomposed tensors, and then fills in the values associated with the combinations of the components of the modes. De Lathauwer [3] presents several ALS methods for computing the HOSVD. Kroonenberg and Jan de Leeuw [10] propose an ALS solution to (3.2) so that at each step an entire set $\{u_i^{(j)}\}_{i=1}^{m_j}$ is solved for some j while everything else is fixed. In other words, the method concentrates on one subspace at a time.

5. Approximation of a Tensor. The well-known Eckart-Young approximation theorem [5, 8] says that if the SVD of a matrix is given by

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, then the best rank-$k$ approximation is given by

$$A_k \equiv \sum_{i=1}^{k} \sigma_i u_i v_i^T.$$

A consequence of this result is that the SVD can be computed via a greedy method which calculates each triplet $\{\sigma_i, u_i, v_i\}$ in sequence. Now we can ask whether or not the Eckart-Young theorem can be extended to tensor rank decompositions; i.e., is the best rank-k approximation of a tensor given by the sum of the first k terms in its rank decomposition? This relates directly to whether or not the greedy orthogonal, strong orthogonal, or combinatorial decompositions produce a corresponding rank decomposition.

In the case of the strong orthogonal rank decomposition, the answer is definitely no, contrary to the result stated in [13], as the following counterexample shows.

Example 5.1. Consider the strong orthogonal rank decomposition of a matrix $A \in \mathcal{T}(m, m, m)$ defined by

$$A = \sum_{i=1}^{6} \sigma_i U_i,$$

where the $\{\sigma_i, U_i\}$ pairs are defined as follows. Let the vectors $a, b, c, d \in \mathbb{R}^m$ be two-by-two orthogonal; then let

$$\begin{array}{lllllll}
\sigma_1 & = & 1.00, & U_1 & = & a \otimes a \otimes a, \\
\sigma_2 & = & 0.75, & U_2 & = & b \otimes b \otimes b, \\
\sigma_3 & = & 0.70, & U_3 & = & a \otimes c \otimes d, \\
\sigma_4 & = & 0.70, & U_4 & = & a \otimes d \otimes c, \\
\sigma_5 & = & 0.65, & U_5 & = & b \otimes c \otimes d, \\
\sigma_6 & = & 0.65, & U_6 & = & b \otimes d \otimes c.
\end{array}$$

Note that $\sigma_3 U_3$ and $\sigma_5 U_5$ can be combined to form the decomposed tensor

$$\gamma_1 V_1 \equiv \sqrt{\sigma_3^2 + \sigma_5^2}\, \frac{\sigma_3 a + \sigma_5 b}{\sqrt{\sigma_3^2 + \sigma_5^2}} \otimes c \otimes d. \tag{5.1}$$

Similarly, $\sigma_4 U_4$ and $\sigma_6 U_6$ can be combined to form

$$\gamma_2 V_2 \equiv \sqrt{\sigma_4^2 + \sigma_6^2}\, \frac{\sigma_4 a + \sigma_6 b}{\sqrt{\sigma_4^2 + \sigma_6^2}} \otimes d \otimes c. \tag{5.2}$$

But,

$$\gamma_1 = \gamma_2 \approx 0.9552 < \sigma_1 = 1,$$

so neither (5.1) nor (5.2) is the best rank one approximation to A; $A_1 \equiv \sigma_1 U_1$ is. However, the best strong orthogonal rank two approximation is given by

$$A_2 \equiv \gamma_1 V_1 + \gamma_2 V_2,$$

because $V_1 \perp_s V_2$ and

$$\gamma_1^2 + \gamma_2^2 = 1.825 > \sigma_1^2 + \sigma_2^2 = 1.5625.$$

Thus, we have a counterexample to any Eckart-Young type theorem for strong orthogonal rank decompositions. □

Example 5.1 can be reworked as follows to show that the combinatorial orthogonal rank decomposition does not yield a best rank-k approximation either.

Example 5.2. Consider the tensor defined in Example 5.1. Let e and f be any vectors that are orthogonal to each other and also to a and b. We can express a combinatorial orthogonal rank decomposition of A as follows.

$$A = \sum_{i_1=1}^{4} \sum_{i_2=1}^{4} \sum_{i_3=1}^{4} \bar{\sigma}_{i_1 i_2 i_3} \bar{u}_{i_1}^{(1)} \otimes \bar{u}_{i_2}^{(2)} \otimes \bar{u}_{i_3}^{(3)},$$

where

$$\begin{array}{lclclcl} \bar{U}_1 & = & a \otimes a \otimes a, & \quad & \bar{U}_3 & = & e \otimes c \otimes d, \\ \bar{U}_2 & = & b \otimes b \otimes b, & & \bar{U}_4 & = & f \otimes d \otimes c, \end{array}$$

and the only non-zero $\bar{\sigma}$'s are

$$\bar{\sigma}_{111} = \sigma_1, \ \ \bar{\sigma}_{222} = \sigma_2, \ \ \bar{\sigma}_{133} = \sigma_3, \ \ \bar{\sigma}_{233} = \sigma_4, \ \ \bar{\sigma}_{144} = \sigma_5, \ \ \bar{\sigma}_{244} = \sigma_6.$$

So, $\mathrm{rank}_{\perp_t}(A) = 4$. The best combinatorial orthogonal rank-1 approximation to A is $\bar{A}_1 = \bar{\sigma}_{111} \bar{U}_1 = \sigma_1 U_1$ (the same as the best strong orthogonal rank-1 approximation). But, the best combinatorial orthogonal rank-2 approximation is yielded by

$$\bar{A}_2 = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} \bar{\gamma}_{i_1 i_2 i_3} \bar{v}_{i_1}^{(1)} \otimes \bar{v}_{i_2}^{(2)} \otimes \bar{v}_{i_3}^{(3)}.$$

Here

$$\bar{V}_1 \equiv V_1 \quad \text{and} \quad \bar{V}_2 \equiv g \otimes d \otimes c,$$

where g is some vector orthogonal to $v_1^{(1)}$, and the only nonzero $\bar{\gamma}$'s are $\bar{\gamma}_{111} = \gamma_1$ and $\bar{\gamma}_{122} = \gamma_2$. □

The problem of whether or not the Eckart-Young result can be extended to the orthogonal decomposition is still an open question. Example 2.3 shows that it is possible to add an orthogonal decomposed tensor to a sum without increasing its rank ($U_1 + U_2$ has rank 2 as does $U_1 + U_2 + U_3$). This is contrary to a fundamental assumption used in the proof of Theorem 2 in [13]. We also have the problem of uniqueness since, by Example 3.4, we know that the orthogonal decomposition is not unique. One possible solution to this problem is the definition proposed on p. 8. We now seek either a proof or counterexample of the following.

Open Problem 5.3 (Eckart-Young extended). *Let the unique orthogonal rank decomposition of a tensor A be given as in (3.1) and assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then the best orthogonal rank p ($p < r$) approximation to A satisfies*

$$\min_{\operatorname{rank}_\perp A_p = p} \|A - A_p\|^2 = \sum_{i=p+1}^{r} \sigma_i^2$$

and is given by

$$A_p \equiv \sum_{i=1}^{p} \sigma_i u_i.$$

6. Conclusions. There are multiple ways to orthogonally decompose tensors, depending both on the definition of orthogonality as well as on the definitions of decomposition and rank. An Eckart-Young type of best rank-k approximation theorem for tensors continues to elude our investigations but can perhaps eventually be attained by using a different norm or yet other definitions of orthogonality and rank.

Computing an orthogonal tensor decomposition is a challenge as well. Most methods are variations on ALS, a method which can be very slow to converge, although recently several authors (c.f., [3, 15]) have presented new ideas.

Acknowledgements. The author is grateful to the referees for their careful reading of this paper and for their many useful comments, including pointing out references [4, 6]. The author also wishes to thank Dianne O'Leary and Paul Boggs for their helpful suggestions in the preparation of this manuscript.

REFERENCES

[1] J. D. Carroll and J.-J. Chang, *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition*, Psychometrika, 35 (1970), pp. 283–319.
[2] R. Coppi and S. Bolasco, eds., *Multiway Data Analysis*, Elsevier Science Publishers, North-Holland, 1989.
[3] L. de Lathauwer, *Signal Processing Based on Multilinear Algebra*, PhD thesis, Katholieke Universiteit Leuven, Belgium, 1997.
[4] J. B. Denis and T. Dhorne, *Orthogonal tensor decomposition of 3-way tables*, in Coppi and Bolasco [2], pp. 31–37.
[5] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.
[6] A. Franc, *Multiway arrays: Some algebraic remarks*, in Coppi and Bolasco [2], pp. 19–29.
[7] A. Franc, *Etude Algébrique des Multitableaux: Apports de l'Algébre Tensorielle*, Thèse de Doctorat, Spécialité Statistiques, Univ. de Montpellier II, Montpellier, 1992.
[8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 3rd ed., 1996.
[9] A. Kapteyn, H. Neudecker, and T. Wansbeek, *An approach to n-mode components analysis*, Psychometrika, 51 (1986), pp. 269–275.
[10] P. M. Kroonenberg and J. de Leeuw, *Principal component analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 45 (1980), pp. 69–97.
[11] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
[12] J. B. Kruskal, *Rank, decomposition, and uniqueness for 3-way and n-way arrays*, in Coppi and Bolasco [2], pp. 7–18.

[13] D. Leibovici and R. Sabatier, *A singular value decomposition of a k-way array for principal component analysis of multiway data, PTA-k*, Linear Algebra Appl., 269 (1998), pp. 307–329.

[14] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.

[15] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors.* In review, 2000.